Name:____ Date: _____________

A2CC: Sample Proportions

Are cell phones a distraction?

A study released by the Centre for Economic Performance at the London School of Economics looked at 91 schools in four cities in England, where 90 percent of teenagers own a mobile phone. The study found that test scores were 6.41 percent higher where cellphone use is prohibited. In another study, more than half of students said they were distracted when other students used their devices.

Suppose we wanted to find the proportion of Roslyn high school students who believe that cell phones, Ipads, and other digital devices are a distraction in the classroom. Do I need to ask EVERY student to get an accurate measure? What if the population of the school was so large that it would be nearly impossible to ask each and every student?

Now, suppose that the true proportion of Roslyn high school students who believe that digital technology is a distraction in the classroom is 70%. That is, $p = .7$. Let's also assume that we don't know this value but would like to estimate it. When we take a sample from the population and find the proportion of the sample, $\hat{p}$, that believe technology is a distraction, is this a reasonable estimate for the true value of p?

Statistics can answer this question for us. But, first, we need to understand how statistics behave.

The statistic, $\hat{p}$. (read as "p-hat")

Suppose Mary takes a random sample of 50 Roslyn students and calculates $\hat{p}$, the proportion of HER sample that finds technology a distraction. Is it likely that her value is $\hat{p} = .7$?

If Bill completes his own survey of 50 students, should we expect his value of $\hat{p}$ to be the same as Mary's?

Suppose it were possible to take EVERY possible sample of size 50 and compute $\hat{p}$ for each one. All of these $\hat{p}'s$ would create their own statistical distribution. It has been proven in advanced statistics that the distribution of $\hat{p}'s$ is **NORMAL** with a mean, $\mu_{\hat{p}} = p$, and a standard deviation, $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$.

(The discussion is not quite that simple and there are some other conditions that we have to verify, but we will leave that discussion for AP Statistics.)

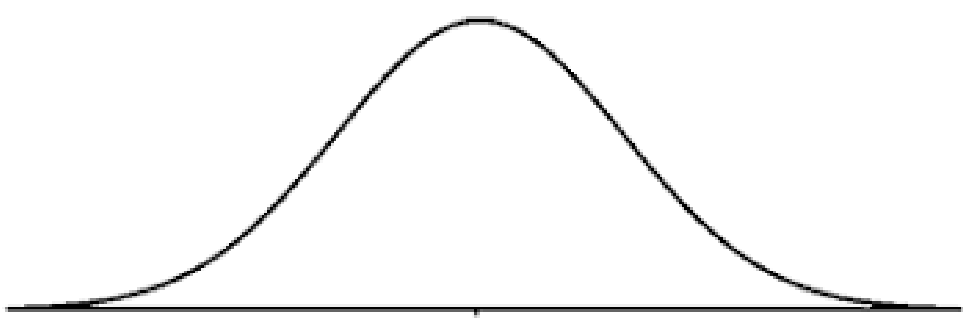
Exercise # 1: Mary takes a random sample of 50 Roslyn high school students. Suppose the true value of $p = .7$ (but remains unknown). What is the probability that Mary finds that at least 76% of her sample think technology is a distraction?

Exercise # 2: Bill takes a random sample of 50 Roslyn high school students. Suppose the true value of $p = .7$ (but remains unknown). What is the probability that Bill finds that less than half of his sample think technology is a distraction?

In both of the previous examples, it was assumed that the true value of $p = .7$. However, if we KNEW that $p = .7$, we wouldn't need to gather data in the first place. This is where statistics come in handy. Since we know the characteristics of the distribution of $\hat{p}$, we can use these characteristics to make inferences about the true UNKNOWN value of p.

Before we can do this, we need to consider one more idea.

Suppose we want to "target" the middle 95% of a normal distribution, how many standard deviations below and above the mean must we go?



Confidence Intervals

So, the middle 95% of the distribution of $\hat{p}'s$ would be represented by $p \pm 1.96\sqrt{\frac{p(1-p)}{n}}$. However, in practice, we don't know p. Since $\hat{p}$ will be close to p in large samples, we replace p in the formula with $\hat{p}$ to create $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. What we've created is a 95% confidence interval for p. We use the term "95% confident" because if we were to construct a confidence interval for each $\hat{p}$ from EVERY possible sample, 95% of the confidence intervals created would contain the true value of p. What does the term 95% confident NOT mean? Most novices think that there is a 95% CHANCE that p is in the confidence interval. This is an INCORRECT interpretation of 95% confidence. The value of p is either in the interval or not. The probability that p is in the interval is either 1 or 0 not .95. We use the term 95% confident to describe the PROCEDURE that was used to create the interval.

Exercise #3: Revisiting Mary's sample from exercise #1. Using her sample, Mary calculated $\hat{p}=.76$. Create a 95% confidence interval for p.

So, we say …

Every confidence interval is of the form: estimate $\pm$ margin of error.

Other common confidence intervals…

A 90% confidence interval for p is $\hat{p} \pm 1.645\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

A 99% confidence interval for p is $\hat{p} \pm 2.576\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

The most commonly used is the 95% confidence interval.

Exercise # 4: The Census Bureau reports that 40% of the 50,000 families in a particular region have more than one smart television in their household. What is the probability that a random sample of size 100 will have at least 45 % of the households that contain a smart television?

Exercise # 5: A large high school has approximately 1200 seniors. The school administration claims that 87% of its graduates are accepted into colleges. If a random sample of 80 graduates is taken, what is the probability that at most 64 of them are accepted into colleges?

Exercise # 6: Suppose that a random sample of 100 high school seniors in a particular city is taken, and it is found that 15% of the students favor the ban on prayer in public schools. Someone argues that the true proportion of seniors that favor the ban is 25%. What do your findings say about this?